AAMRL-TR-89-036

ACOUSTIC CHARACTERISTICS OF SENTENCES PRODUCED IN NOISE

Z. S. Bond
Thomas J. Moore
Kate McCreight

BIODYNAMICS AND BIOENGINEERING DIVISION
HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-6573

SEPTEMBER 1989

SUMMARY REPORT FOR AUGUST 1988 TO SEPTEMBER 1989

Approved for public release; distribution is unlimited.

HARRY G. ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
HUMAN SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Harry G. Armstrong Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield VA 22314

TECHNICAL REVIEW AND APPROVAL

AAMRL-TR-89-036

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Dennis A. Reed

DENNIS A. REED, Lt Col, USAF, BSC
Associate Director
Biodynamics and Bioengineering Division
Harry G. Armstrong Aerospace Medical Research Laboratory

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AAMRL-TR-89-036	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Harry G. Armstrong Aerospace Medical Research Laboratory 6b. OFFICE SYMBOL (If applicable) AAMRL/BBA	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Wright-Patterson AFB OH 45433-6573	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F PROJECT NO. 2313 TASK NO V3 WORK UNIT ACCESSION NO. 01

11. TITLE (Include Security Classification)
(U) ACOUSTIC CHARACTERISTICS OF SENTENCES PRODUCED IN NOISE

12. PERSONAL AUTHOR(S)
Z. S. Bond, Thomas J. Moore and Kate McCreight

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Summary	FROM Aug 88 TO Aug 89	89 Sep 1	23

16. SUPPLEMENTARY NOTATION
AAMRL/Contact: Dr Thomas J. Moore, AAMRL/BBA, Tel: (513) 255-3607

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
25	04	
01	08	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Acoustic phonetics
Voice communications
Noise effects

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Previous work in a number of laboratories has described relatively systematic changes in the acoustic-phonetic structure of speech produced in the presence of noise relative to that produced under more benign speaking circumstances. The purpose of this study was to determine whether noise affects the production of continuous speech as it does words produced in isolation. Four speakers were recorded reading 20 sentences, two times in quiet and two times while having 95 dB SPL of pink noise presented through headphones. The sentences were digitized, segmented and transcribed using SPIRE. The resulting data base consisted of approximately 850 segments per subject per speaking condition. The distributions of spectral and temporal properties of classes of segments were determined using SEARCH. All sentences produced in the presence of noise had increases in fundamental frequency and total energy, as has been found for isolated words. Segment durations and spectral characteristics were affected by noise for some subjects more than for others.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL T. J. MOORE	22b. TELEPHONE (Include Area Code) (513) 255-3607 22c. OFFICE SYMBOL AAMRL/BBA

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

PREFACE

This research was accomplished in the Biological Acoustics Branch, Biodynamics and Bioengineering Division, Harry G. Armstrong Aerospace Medical Research Laboratory, Human Systems Division (HSD). The effort was accomplished under Work Unit 2313V301, "Auditory Information Processing."

The research was sponsored by the Air Force Office of Scientific Research/AFSC, United States Air Force.

The first author is a faculty member at Ohio University, Athens, OH; the third author is a faculty member at Wright-State University, Dayton, OH.

The authors wish to acknowledge the aid of Mrs Hazel Watkins in the typing and preparation of this report.

TABLE OF CONTENTS

INTRODUCTION

Since the work of Lombard (1911), we have known that when speakers talk in the presence of noise, characteristics of their speech change. Recently, there has been considerable interest in describing the details of these acoustic-phonetic changes. Summers, et al., (1988) reported that amplitude, fundamental frequency, and segment durations increased in the presence of noise. In addition, they found differences in formant frequencies and the short-term spectra of vowels. Such changes were also described by Bond, Moore, and Gable (1989), though we reported some subject variability in the effects of noise on segment durations.

The purpose of this study was to extend our understanding of the effects of noise on speech by examining sentences rather than isolated words produced while speaking in the presence of a relative high level of noise. It is known that the global effects of increases in fundamental frequency and amplitude found in isolated words are also found in continuous speech produced in noise environments (see Lane and Tranel, 1971). What is not known is whether the segmental and spectral effects observed in isolated words are also present in connected or continuous speech.

METHOD

Speakers

The speakers were four young males, college students at a Midwestern university. None of the speakers had any history of speech or hearing difficulties. All were audiometrically screened to ensure that they had Hearing Threshold Levels of less than 15 dB. They also served as listeners on a panel investigating speech intelligibility for the Air Force and consequently were experienced speaking in noise environments. These same four speakers served in an earlier study (Bond, et al., 1989). The subjects were paid for their participation.

Procedure

The speakers were recorded in a baseline condition with no noise exposure and while listening to pink noise over headphones at 95 dB SPL. Both recordings were made using a military boom microphone (M-167) while the subjects were seated in an anechoic chamber. Side tone was adjusted by the speakers to what they considered a comfortable level in the baseline condition and was not changed when the speakers were exposed to noise.

The speakers recorded 20 short sentences, taken from the CID sentence lists (lists E & J, Davis and Silverman, 1978), 2 times

in each speaking condition, for a total of 80 sentences per subject. The speakers read the sentences in a relaxed, relatively casual speaking style.

Data Analysis

Speech analysis was performed using SPIRE (Speech and Phonetics Interactive Research Environment), on the Symbolics 3670 computer. Each production of each sentence was digitized at 16 kHz with 16 bit resolution. Each segment in each sentence was labeled using the transcription facility of SPIRE (Cyphers, et al., 1986). Segment boundaries were located from wide-band spectrogram and waveform displays, following the criteria outlined in Peterson and Lehiste, 1960. Word boundaries were also marked. The data set consisted of approximately 850 labelled segments for each speaker in each speaking condition.

The SPIRE parameters of formant frequencies, fundamental frequency, frication frequency, total energy, and energy in low and high frequency bands were computed for all segments in each speaking condition for each subject. These samples were submitted to the program SEARCH (also developed by the Speech Processing Group at MIT) so that speech parameters of interest could be compared in both speaking conditions for any segment or group of segments.

SEARCH allows data sets describing utterances to be partitioned into user-specified subsets, for example all stops, or all voiceless fricatives. SEARCH also calculates simple descriptive statistics of SPIRE parameters for phoneme subsets, e.g., means and standard deviations of the duration of all fricatives or the frequency of the first formant for all vowels. (See Cyphers, et al., 1986, for further details).

RESULTS

Fundamental Frequency

As in almost all previous investigations, the read sentences were found to be higher in pitch when the speakers were exposed to noise than when they were speaking in the benign condition. The fundamental frequency, taken at the mid-point of all vowels in the sample, increased for each of the four speakers in noise. The distributions of the fundamental frequencies are given in Fig. 1 for each speaker. The smallest average fundamental frequency (Fo) increase was 13 Hz for S4, the greatest was 48 Hz for S2. Averaged for all four speakers, Fo increased 25 Hz, approximately a 26 percent increase. There was also a tendency for the variability of Fo to increase for speech produced in the presence of noise.

Total Energy

Total energy also increased for all four speakers in the presence of noise. Total energy values per speaker, averaged for all vowels in the sample, are given in Fig. 2. Total energy is measured using SPIRE in terms of dB down from a reference. The largest total energy increase, 11 dB, was found for S2, the speaker who also exhibited the greatest increase in fundamental frequency. Averaged for four speakers, the total energy increase was 7 dB. In general, increases in total energy and fundamental frequency were correlated. Increases in total energy were associated not only with vowels but with all other segments for which energy could be measured.

Spectral Tilt

The spectrum of speech produced in noise has also been found to be characterized by a relative increase in energy in high frequencies in comparison with lower frequencies, that is, by a change in spectral tilt. In order to evaluate the read sentences for this possibility, the energy in a low-frequency band (300-600 Hz) and a high-frequency band (2000-3000 Hz) was calculated for all vowels. Since total energy increased with noise, energy would be expected to increase in both energy bands as well. The increase in the low-frequency band averaged 6.9 dB for all four speakers while the energy in the high-frequency band increased almost 10 dB. For all four speakers, there was a tendency for more energy to be present at higher frequencies for speech produced in noise.

Durations

The overall noise effects on word and segment durations in read sentences were variable for the four subjects. For two subjects, the average durations of all words decreased in noise, by 41 ms for S1 and 14 ms for S3. For the other two speakers, average word durations increased by 18 ms for S4 and by 5 ms for S2.

For three speakers (S2, S3, S4) the average durations of all vowels increased by a very small amount, from 3 to 15 ms. For S1, average vowel durations decreased by 15 ms. The tendencies found for all vowels were also present for vowel subsets such as inherently long and short vowels and diphthongs. In general, the longer the vowel, the more it tended to increase in duration. The magnitude of the effect of noise on vowel durations, however, was clearly small and statistically non-significant. The distributions of vowel durations for all four subjects are given in Fig. 3.

Frication Frequency

In SPIRE, frication frequency is defined as the most prominent frequency in noisy portions of the speech signal. Averaged across all fricatives, frication frequency increases for all subjects by 370 Hz, or approximately 18 percent. The values for each speaker are shown in Table 1.

Vowel Formants

Values for the first and third formants averaged across all vowels for each speaker are given in Table 2. The most consistently reported effect of noise on the formant structure of vowels has been an increase in the frequency of the first formant. This effect was present and can be seen both for individual vowels and globally. Averaged for all vowels in the sample, the first formant increased from a maximum of 71 Hz (S2) to a minimum of 10 Hz (S3). When averaged for all four subjects, the first formant increased 34 Hz.

The second most consistent vowel formant shift affected the third formant. On the average, the third formant was lower in speech produced in noise for all four subjects. Averaged for all vowels, the third formant decreased by 140 Hz for S1 to 50 Hz for S3. The average for all four subjects was a decrease of 88 Hz.

Second formant values averaged across all vowels are not reported because previous work suggests that the effects of noise on the second formant may vary from vowel to vowel. (Bond, et al., 1989).

Figure 4 shows the average center frequencies of F1 and F2 for the four vowels /i, ae, a, u/, which represent the corners of the traditional vowel quadrilateral, produced under both ambient

and noise conditions. As has been reported for isolated words, the major effect of speaking in the presence of 95 dB pink noise is an upward shift in frequency of F1. As we also observed in the case of isolated words, F2 for /i/ shows a slight decrease in frequency while it remains essentially unchanged for /ae/ and /a/. The major difference between the results noted in the vowel F1-F2 plots for sentences and those reported for isolated words occurred with /u/. In the isolated word condition words spoken by the same four talkers resulted in an upward shift of F2 for /u/ when spoken in the presence of noise; in the sentences F2 for /u/ decreased slightly when spoken in noise relative to the ambient condition. The major difference, however, was a significant increase in F2 for /u/ when embedded in a sentence as opposed to when in an isolated word. When in isolated words the average F2 value for /u/ produced by the four talkers in ambient conditions was about 1000 Hz. When the same four talkers under the same conditions read sentences, the average F2 value for /u/ was around 1650 Hz. Fokes and Bond (1986) have noted that there is a tendency for American talkers to produce /u/ with a higher second formant in sentence context then when the same vowel appears in isolated words. However, the difference they noted was not as pronounced as that found here.

DISCUSSION

The changes of speech with noise observed in sentences are consistent with our previous findings dealing with isolated words

and also with the general tendencies reported in the literature. First, duration changes for words and segments are small and inconsistently present. They do not appear to be systematic enough to attribute to the noise environment, though possibly S1 is an exception.

Second, increases in pitch frequency and total energy as well as in frication frequency are present for all speakers. These changes probably result from increased vocal effort. When in the noisy environment, the speakers try to increase the loudness of their speech to a level they feel appropriate. The changes in spectral tilt would be an expected consequence of increased vocal effort as well.

Third, the formant changes are also generally consistent with previous work. The increase of F1 may be a consequence of restricted tongue movement caused by the more open mouth position associated with loud speech. However, an explanation for the systematic decrease in F3 is not entirely clear. A low F3 is associated with a mid-palatal constriction at least in the production of rhotacized vowels (Pickett, 1980). Whether a palatal constriction is responsible for the observed F3 decreases or whether they result from some other speech production mechanism, perhaps pharyngeal stiffening, is not clear on the basis of this research. That pharyngeal stiffening may be responsible for the F3 shift is suggested by a finding of Butcher and Ahmad (1987), who report a lowering of F3 by approximately

200 Hz in the environment of the pharyngeal consonants of Iraqi Arabic.

Finally, it has been noted (Bond, et al., 1989; Moore and Bond, 1987; Summers, et al., 1988) that many of the changes observed in speech produced in noise may reflect articulatory changes made to increase vocal effort and to more precisely articulate in order to enhance communication in an interfering environment. Indeed it has been shown that for equivalent signal-to-noise ratios, speech produced in noise is more intelligible than speech produced in quiet (Dreher and O'Neill, 1957; Summers, et al., 1988). In addition, we have conducted listening tests using the isolated words spoken by these same four talkers (Bond and Moore, 1989) and found that the words produced in noise were more intelligible at equivalent signal-to-noise levels for both native and non-native speakers of English, with the non-native speakers of English showing the greater increase in intelligibility.

REFERENCES

1. Bond, Z. S., Moore, T. J. and Gable, B. (1989). Some acoustic phonetic characteristics of speech produced in noise and wearing an oxygen mask, J. Acoust. Soc. Amer., 85, 907-912.

2. Bond, Z. S. and Moore, T. J. (1989). "Intelligibility of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Amer., 85, Suppl 1, S55.

3. Butcher, A. and Ahmad, K. (1987). "Some acoustic and aerodynamic characteristics of pharyngeal consonants in Iraqi Arabic," Phonetica, 44, 156-172.

4. Cyphers, D. S., Kassel, R. H., Kaufman, D. H., Leung, H. C., Randolph, M. A., Seneff, S., Unverferth, J. E., Wilson, T., and Zue, V. W. (1986). The Development of Speech Research Tools on MIT's Lisp Machine-Based Workstations, Speech Recognition: Proceedings of a Workshop, Palo Alto, CA, Science Applications International Corp. Report No. SAIC-86/1546, 110-115.1

5. Davis, H. and Silverman, S. R. (1978). Hearing and Deafness (4th Ed.), NY: Holt, Rinehart.

6. Dreher, J. J. and O'Neill, J. J. (1957). "Effects of ambient noise on speaker intelligibility for words and phrases," J. Acoust. Soc. Amer., 29, 1320-1323.

7. Fokes, J. and Bond, Z. S. (1986). "Vowel quality and word stress in native and non-native English," J. Acoust. Soc. Amer., 80, Suppl. 1, S50 (A).

8. Lane, H. L. and Tranel, B. (1971). "The Lombard sign and the role of hearing in speech," J. Sp. Hring Res., 14, 677-709.

9. Lombard, E. (1911). "Le signe de l'elevation de la voix," Ann, Mal. Oreil. Larynx, 37, 101-119. (Cited by Lane and Tranel, 1971).

10. Moore, T. J. and Bond Z. S. (1987). "Acoustic-phonetic changes in speech due to environmental stressors: Implications for speech recognition in the cockpit." Proceedings of 4th Aviation Psychology Symposium, Columbus, Ohio, 77-83.

11. Peterson, G. and I. Lehiste (1960). Duration of syllable nuclei in English, J. Acoust. Soc. Amer., 32 693-703.

12. Pickett, J., (1980). The Sounds of Speech Communication. Baltimore: University Park Press.

13. Summers, V. W., Pisoni, D. B., Pedlow, R. I. and Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses, J. Acoust. Soc. Amer., 84, 917-928.

TABLE 1. FRICATION FREQUENCY (Hz)

SUBJECT	AMB.	NOISE	CHANGE
1	1820	2130	310
2	1930	2250	320
3	2070	2460	390
4	2310	2770	460
Average	2032.5	2402.5	370

TABLE 2.

SUBJECT	F_1 (Hz) AMB.	NOISE	CHANGE	F_3 (Hz) AMB.	NOISE	CHANGE
1	447	473	26	2330	2190	-140
2	448	519	71	2390	2290	-100
3	433	443	10	2380	2330	- 50
4	506	533	27	2540	2480	- 60
Average	458.5	492	33.5	2410	2322.5	-87.5

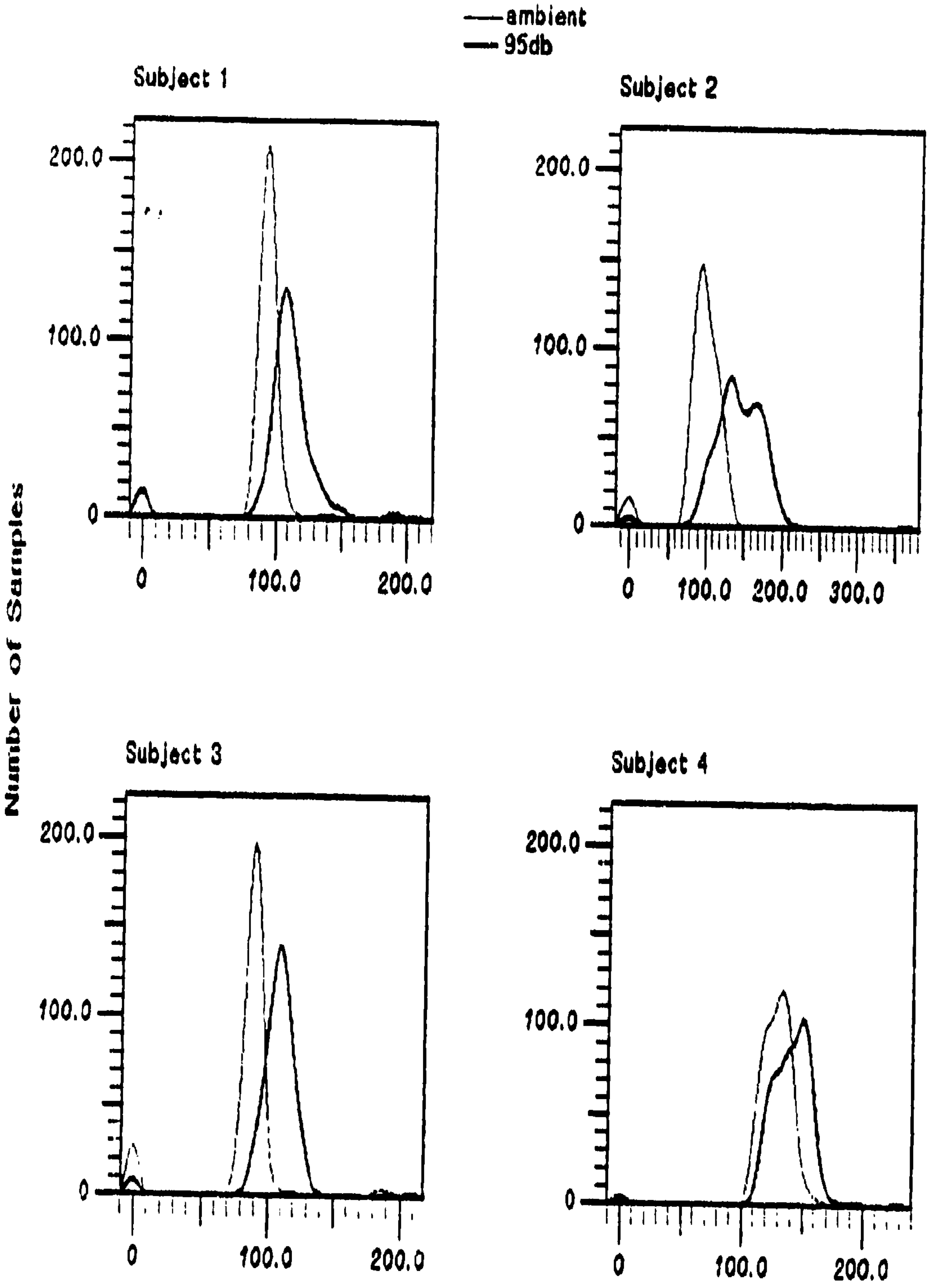

Fig. 1. Distribution of pitch frequency in both speaking conditions for four speakers. The abscissa is in Hz.

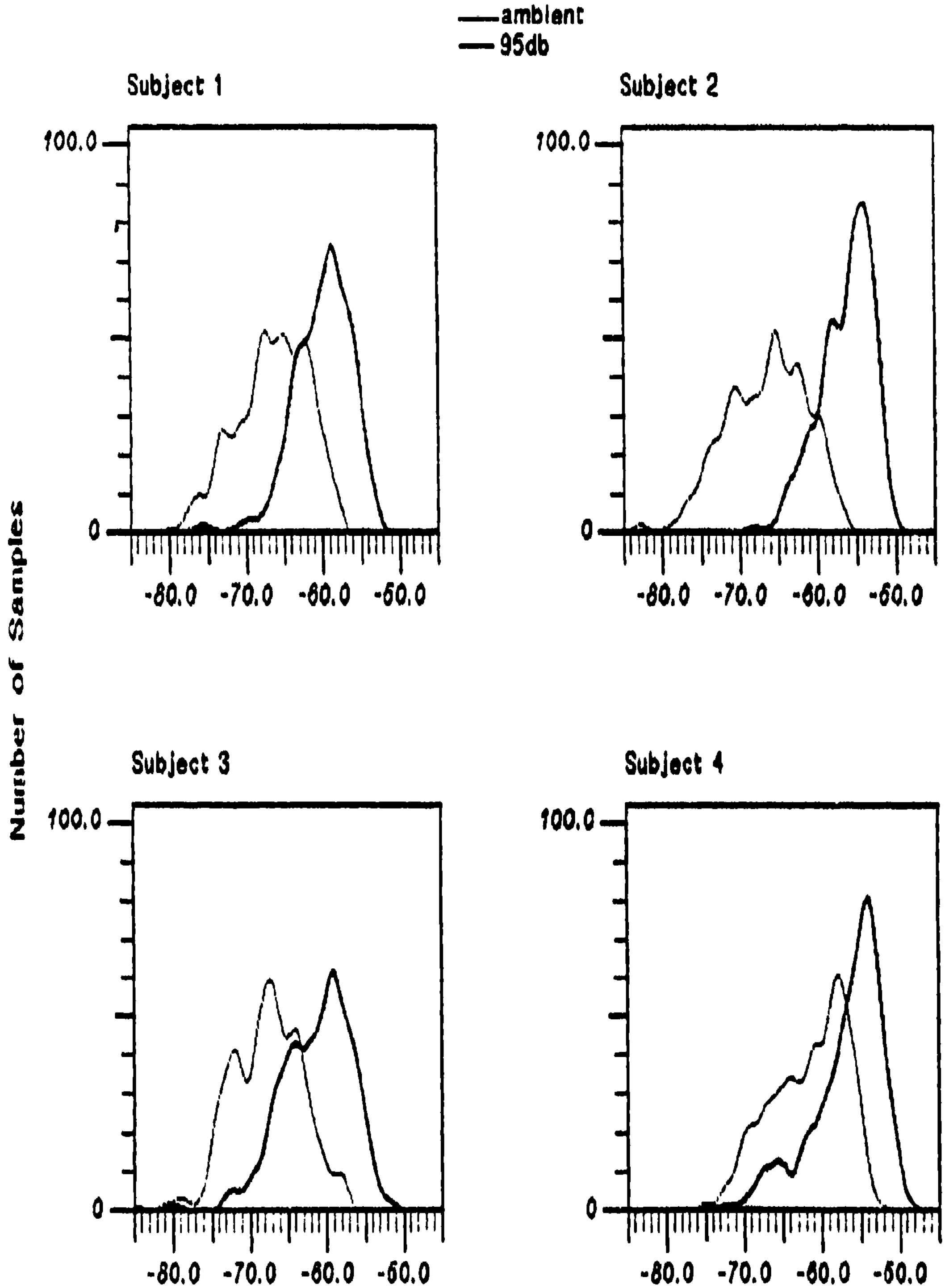

Fig. 2. Total energy in both speaking conditions, four speakers. The abscissa is in dB down from a reference level.

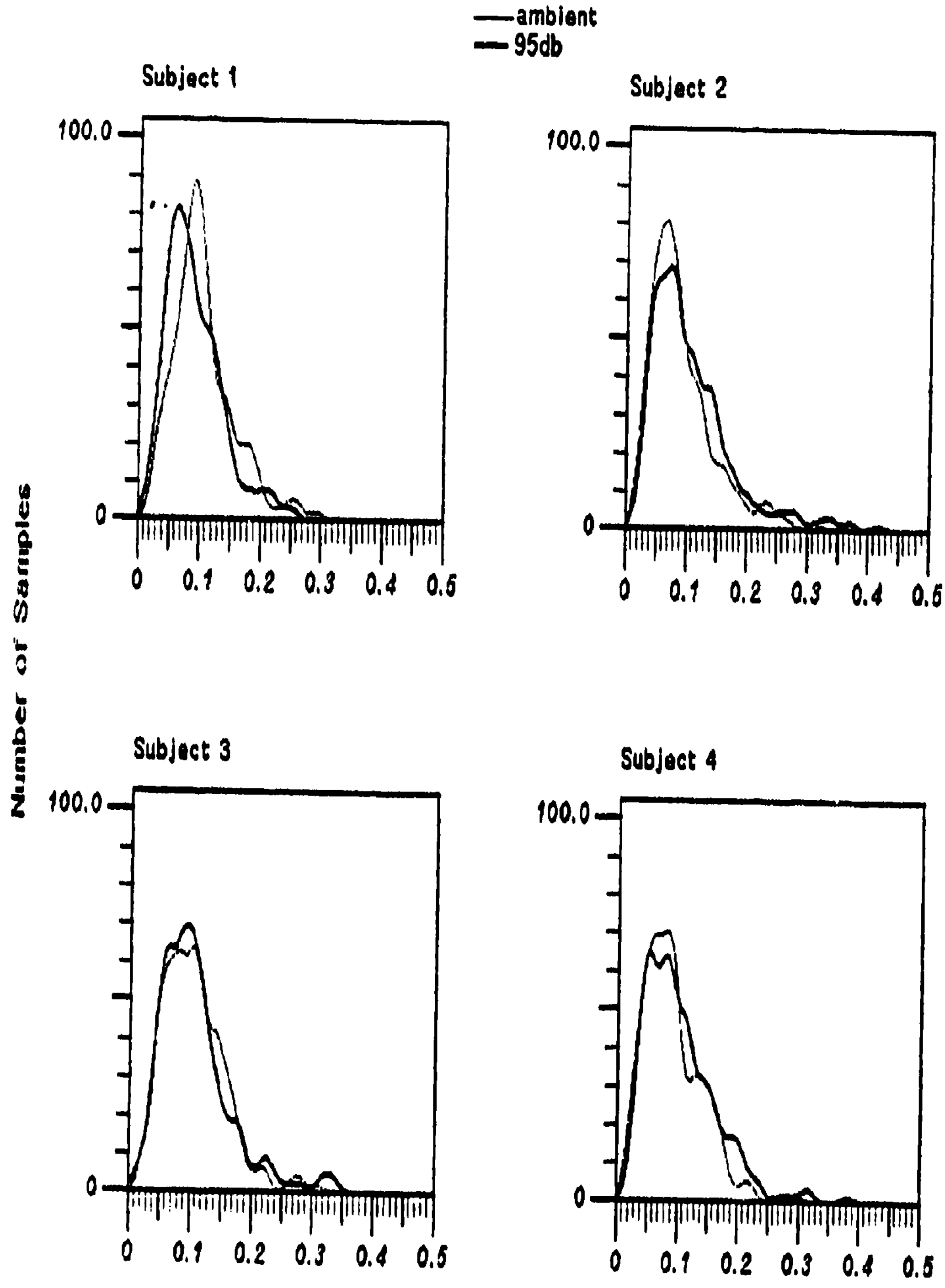

Fig. 3. Distribution of vowel durations in both speaking conditions for four speakers. The abscissa is in seconds.

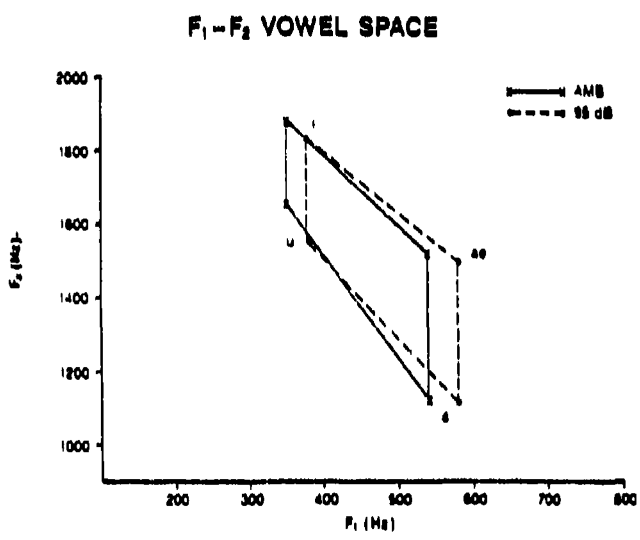

Fig. 4. The space defined by F1-F2 for the front vowels /i ae/ and for the back vowels /u,a/, in two conditions.

★U.S. Government Printing Office: 1989 — 748-002/00216